Transportation 13: 271–293 (1986)

Classification issues in the analysis of complex travel behavior

SUSAN HANSON[1] & JAMES HUFF[2]
[1]*School of Geography, Clark University, Worcester, MA 01610;* [2]*University of Illinois, Urbana, Ill.; USA*

Abstract. Researchers have devoted considerable effort to identifying homogeneous travel-behavior groups, each of which also has distinctive sociodemographic characteristics. Interest in these efforts has been fueled by both theoretical and applied concerns. From a theoretical perspective, if such behaviorally *and* sociodemographically homogeneous groups can be identified, we would have an improved understanding of the determinants of travel. From an applied perspective, because spatial choice models assume behavioral homogeneity within each sociodemographically defined subgroup, one would like to be able to identify groups that are homogeneous with respect to both behavior and sociodemographics. Most previous efforts to define such groups have classified individuals on the basis of one-day travel records. In this paper we review these efforts, note the problems inherent in using one-day travel records for identifying homogeneous travel-behavior groups, and use standard grouping procedures to classify individuals on the basis of behavior observed over a longer time period (five weeks). Using multi-day travel data means that the travel measures employed in classifying individuals are different from and more complex than those used with one-day data. We identify five travel-behavior groups, each of which has distinctive sociodemographic characteristics. Considerable intra-group variability remains, however, even though the groups are classified on the basis of longer-term behavior. The paper concludes with an examination of the implications of day-to-day variability in individual travel for classification procedures.

The quest for an understanding of the mechanisms generating urban travel has recently taken the path of attempting to define groups of individuals with similar travel-activity patterns and to identify variables (usually descriptors of the sociodemographic and locational characteristics of those individuals) that can successfully account for the between-group variance in such patterns. The choice of this particular path seems logical; it reflects the belief that depicting the patterns and the correlates of the patterns is the first step in comprehending travel-activity patterns, in grasping how and why they change, and in predicting future patterns successfully. Perhaps it is the scientist's conviction that the world operates in a fundamentally orderly fashion that convinces travel behavior researchers that there do exist groups of people with essentially similar spatial behavior and that these groups can be distinguished by meaningful independent variables. Certainly the search for such groups is motivated at least in part by the fact that disaggregate choice models, by segmenting the population on sociodemographic descriptors and then calibrating a choice model for each resultant subgroup, assume the existence of homogeneous behavioral/sociodemographic groups. Oppenheim captures the essence of this assumption in the hypothesis central to his own empirical study: "...if two travelers have the same personal characteristics and live in similar urban environments, they will have the same travel behavior" (Oppenheim, 1975, p. 141).

One reason for the plausibility of expecting to find distinguishable travel-activity patterns among different well-defined groups is the intuitively reasonable notion that the longer-term decisions of individuals and households systematically affect their short-term behavior (see for example Salomon, 1983; Cullen & Phelps, 1975). Such long-term decisions as those regarding employment (e.g. the type and amount of paid work to undertake and the location of such work), car ownership, marriage, and residential location are thereby seen as determining (at least to some extent) the daily travel pattern. There is a corrollary to this line of thinking, and it is an assumption that usually remains unarticulated in studies of spatial behavior. This is the belief that the daily travel-activity pattern of any individual or household is largely habitual. The twin assumptions that long-run decisions govern short-term behavior and that behavior varies little in the short run feed a vision of behavior as eminently predictable. Once the larger, longer-term decisions have been made, according to this view, the individual quickly falls into a repetitive daily pattern, the shape of which depends on such things as auto ownership, sex, employment status, marital status, and residential and work locations.

The belief in the existence of orderly, identifiable behavioral/sociodemographic groups is so attractive and tenacious a notion that it persists in the face of evidence that is at best partial and tentative and at worst confusing and conflicting. Whether individuals are grouped first on the basis of their sociodemographic characteristics, such as roles (e.g. Van der Hoorn, 1979; Janelle & Goodchild, 1983), or first on the basis of their daily travel behavior (e.g. Recker & Schuler, 1982; Pas, 1984), the results are similar: groups defined on the basis of sociodemographics are not isomorphic with (in fact they often bear little relation to) groups defined on the basis of travel-activity patterns. And yet there has been some success, albeit limited, in relating the individual's sociodemographic and locational characteristics to his or her travel pattern (e.g. Pas, 1984; Recker & Schuler, 1982; Hanson & Hanson, 1981; Allaman *et al.*, 1982). Existing evidence (reviewed in detail below) is therefore somewhat mixed, indicating as it does that although certain types of individuals do display similar travel-activity patterns, the levels of intragroup variability in measures of travel behavior remain high.

Before the search for order is abandoned as futile, there is an important question that needs scrutiny: to what extent is intragroup variability in behavior patterns an artifact of a particular type of data base? Most studies have defined travel behavior groups on the basis of travel records for only one day. The validity of using a one-day data base to identify homogeneous behavioral groups rests on the assumption that within-group behavior is homogeneous not only in the long run, but also in the (very) short run. That is, individuals within each group are implicitly assumed to exhibit *precisely the same type of habitual behavior.* As Recker & Schuler (1982, p. 118) have recognized, defining groups of individuals with similar behavior on the basis of a one-day travel record is fraught with problems. Depending upon whether a person happened to take (or not to take) a particular type of trip (such as a shopping trip or a social/recreation trip) on the diary day, he or she will be assigned to one group as opposed to another. As long as the travel measures used to classify individuals are fairly detailed, groups defined on the basis of one-day travel records are likely to be highly unstable. Members of different groups in Recker & Schuler's analysis, for example, could actually, as they point out, exhibit quite similar behavior patterns if behavior were observed over some longer time period.

Although habitual behavior has long been assumed the norm and although travel behavior models and theory have been tested almost entirely with one-day data bases, it is no secret that for any individual one day's travel pattern is probably not exactly like every other day's pattern. Curry (1967), for example, posited regular cycles of behavior in his deductive analysis of the relationships between travel and central place systems. A reasonable working hypothesis is that the individual's behavior is neither completely random nor completely habitual but is in fact cyclical – characterized by what might be called systematic variability, with certain daily patterns or fragments therefore recurring at known intervals. Acknowledging day-to-day variability in individual travel throws doubt on the validity of attempts to group people on the basis of one-day trip records. The existence of temporal variability suggests that classification should proceed on the basis of travel behavior recorded over some extended time period rather than for one day. In this paper we first review other work aimed at identifying homogeneous travel behavior groups (each having distinctive sociodemographic attributes) and then seek to identify groups of individuals having similar travel behavior, where the measures are drawn from a five-week record of out-of-home travel-activity patterns.

Background

The assumption that an individual's personal characteristics systematically affect daily behavior stretches back to early time budget studies (e.g. Szalai, 1972; Chapin, 1974) which examined differences among population subgroups in time use. It is an idea that still wields considerable force despite difficulties encountered in assembling the evidence base needed to underpin it.

Two main approaches have dominated attempts to obtain the necessary evidence. The first has been widely used in time budget studies; it consists of defining a priori population subgroups (for example, working men, full-time working women, part-time working women and so on) and then examining the variations in time-use patterns among the different groups. Until the past two or three years most studies focusing on people's use of space, or of space and time, adopted essentially the same approach: they first established sociodemographically-defined groups and then searched for behavioral differences among the groups (see Hanson & Hanson, 1981, for a review of such studies). It is worth noting that an explicit concern for the level of intra-group variability in the behavioral measures was not, until recently, characteristic of the studies that followed this approach. Van der Hoorn's (1979) study is one that reflects the emerging recognition of the problem of intra-group variation. Having set out to find "...more or less homogeneous person categories and to derive some broad macro relations regarding their behavior" (p. 312), Van der Hoorn first divided his sample from the Netherlands into five groups (working men, working women, housewives, students, and others). When he examined travel times and travel frequencies for each of these groups, Van der Hoorn claimed to have found differences between the groups (although no tests of significance are presented), but he was also troubled to find high levels of intragroup variation. Moreover, within each separate group there was no strong relationship between the measures of activity pattern (travel time and travel frequency) and "...socioeconomic variables like age, income, etc." (p. 326).

Most recent studies have adopted the alternative approach of first grouping individuals on the basis of their observed behavior and then searching for sociodemographic or locational characteristics that help to explain such behavioral groupings. Pas (1984), for example, with data from Baltimore identified groups of individuals on the basis of their one day travel-activity patterns. He then sought to determine via regression analysis whether or not each of a number of sociodemographic characteristics had a significant impact on the daily travel-activity pattern and found that a large number of them did. Similarly Recker & Schuler (1982) cluster individuals first on the basis of their one-day travel activity patterns and then use discriminant analysis to examine the impact of a set of sociodemographic and locational descriptors on these travel activity patterns. The authors were able to identify a set of such descriptors that did discriminate among the different groups to some extent. Dangschat *et al.* (1982) adopted a similar approach, but found that sociodemographic descriptors were very poor discriminators among groups of individuals defined on the basis of their travel-activity patterns (measured by participation in work, shopping, and social activities). They found that sociodemographic characteristics were distributed about the same across all five behaviorally defined groups. Similarly Oppenheim's (1975) initial optimism about the similarity between (a) groups defined on the basis of behavior and (b) groups defined on the basis of sociodemographic characteristics turned to sheer pessimisms when he examined his data.

Several researchers have used modified versions of these two approaches, but even grouping people on a mixture of behavioral and sociodemographic variables has not yielded groups that were highly homogeneous with respect to behavior. Herz (1982, 1983) divided his German sample into 271 a priori groups on the basis of sociodemographic and locational characteristics. This large number of groups was then collapsed to 7 by maximizing between-group and minimizing within-group variability with respect to behavior (time spent on different activities, frequency of different activity sequences, time spent travelling by different modes). Despite some intra-group similarities in travel-activity patterns, Herz found a disturbingly high level of intragroup variability: "At the best, 30% of the behavioral variability between individuals is explained by these sociodemographic groupings, and for hourly travel time budgets, broken down by the three main modes, it is only 3%" (Herz, 1983, p. 392). Using a modified version of the a priori group-definition approach, Clarke *et al.* (1981a) have had somewhat better success in explaining the differences between groups defined on the basis of stage in the life cycle. With information gleaned from qualitative interview data about people's daily activity patterns, Clarke *et al.*, identify life cycle groups and an archtypical activity pattern for each group. They then investigate actual time use via one-week activity diaries, and are able to discern significant differences among most of the life cycle groups' patterns of time use.

In general, these recent studies have found that sociodemographic descriptors, and particularly role descriptors, are better discriminators of travel-activity patterns than are locational characteristics; yet sociodemographic characteristics still explain only a relatively small amount of the variation in behavior patterns between individuals (Herz, 1982; Recker & Schuler, 1982; Wermuth, 1982; Kutter, 1973; Pas, 1984; Vidakovic, 1983; Allaman *et al.*, 1982; Hanson, 1982). The degree of success achieved by sociodemographics in explaining travel behavior depends, of course, on the complexity of the behavioral measures used. Sociodemographics (especially stage in the life cycle) are more effective as discriminators when the behavioral measure is the relatively simple one of time allocated to different types of activities (as in Clarke *et al.*, 1981a; and Allaman *et al.*, 1982) and less effective when complex measures of space-time behavior are used (as in Herz, 1982, 1983). When the behavioral measure is the daily activity (or activity program), for example, there is high intragroup variability in the number of different types of daily programs observed. Still the notion persists that "individuals with similar role sets possess similar behavior..." (Dangshat *et al.*, 1982, p. 1167), and modelers continue to erect their models on this assumption. Vidakovic's model for example, assumes that the frequency of visits of a particular type "is entirely based upon [the individual's] socio-occupational and accessibility/mobility situation" (Vidakovic, 1983, p. 179). That is, for all individuals in a given sociodemographic group, the probability of making a particular type of visit is invariant.

With the exception of two studies (Van der Hoorn and Clarke *et al.*), all of the studies reviewed above used behavioral measures derived from a one-day window on travel. The Van der Hoorn study uses a one-week diary, but he recognizes numerous problems with the data, not the least of which were that the diaries record only the major activity for each quarter hour and the fact that the trip destinations and distances traveled are not known.

If indeed relying on a one-day data base makes it difficult to identify groups of individuals with similar behavior, and if repetitive or habitual behavior does exist, then we might aim to classify individuals on the basis of their long-term behavior. This raises the question, however, of how we might measure such behavior. In this context there has been some interest in trying to elicit individuals' usual or habitual behavior from reported estimates on questionnaires. Godard (1983) has, however, documented some of the problems associated with trying to collect data on habitual behaviors in a survey rather than a diary format. By comparing respondents' answers to questions about their habitual behavior with their responses on a one-day diary, he has shown that when people are asked to identify their habitual activities and the frequency of each, they overestimate the frequency with which they participate in most activities, and the amount of overestimation depends on the type of activity in question.

If it is not possible to obtain accurate measures of habitual behavior from questionnaires, is it possible to derive them from one-day records of travel behavior? Even if one selects a mid-week day so as to capture the traveler's "most typical" day (as recommended by Adiv, 1983), it is difficult to know whether that day *is* the most typical. It is also hard to discern habitual behavior from a one-day window because the one-day data base does not permit the researcher to observe much in the way of variability in the behavior pattern. What constitutes habitual behavior cannot really be established with one-day data sources; it can only be assumed. As a consequence, any variation in travel behavior observed at the aggregate level must be assumed to spring from unexplained heterogeneity within the population, whereas much of the observed variance may in fact be due to nonrandom, day-to-day variation in individual travel behavior. In order to classify individuals in a way that accommodates temporal variability, data collected over some period of time longer than one day (for example a week or a month) are required.

Data

This study uses data on out-of-home travel-activity patterns collected in Uppsala, Sweden. The Uppsala Household Travel Survey made use of self-administered travel diaries to collect detailed information on all out-of-home movements by the adult members of a sample of households over a 35-day period. In this paper we use a sample of 149 individuals from 93 households that constitute a representative sample of Uppsala's population. Each home-to-home circuit was defined as a series of movements consisting of one or more stops that began and ended at home. For each stop on each circuit, the panel members recorded the time of arrival at and departure from the stop, the exact location and purpose of the stop, and the mode of travel used to arrive at the stop. Three unusual aspects of this data set are; (1) the inclusion of all non-motorized (walking and bicyle) movements in addition to vehicular movements; (2) the use of an extremely detailed locational coding scheme in which the origin and destination of each movement are associated with street addresses; and (3) the length of time over which behavior is observed. The fact that the data cover a five-week period does limit to daily and weekly cycles the types of repetitive behavior that can be observed; monthly and yearly cycles in travel behavior cannot be captured in a 35-day record. Nevertheless, no other data set on travel behavior covers in detail the entire travel-activity pattern of a sample of individuals for as long as five weeks.

One of the problems with using data collected over so long a time period is the vulnerability to bias if respondents do not record their behavior with persistent precision over time. A commonly held belief among transportation researchers is that respondents will become increasingly lax in recording trips as they loose interest over the course of a longitudinal study. In fact Clarke *et al.* (1981b) report a slight reduction in the number of trips reported per day over a 7-day period in one English study. Therefore, the first question to address to any longitudinal data set is the degree to which recorded travel remains stable over the observation period. We have examined this question in detail elsewhere (Hanson & Huff, 1982) and have found no significant trends in either the number of stops or the number of trips made per individual per day over the 35-day study period. For over 90% of the individuals in our sample both the number of trips made per day and the number of stops made per day were stable; for those few individuals where a trend in the rate of trip making was discernable, the trend was not consistently one of reduced tripmaking over time.

Analysis

The analysis seeks to group individuals on the basis of their observed travel behavior over the 35 days and to characterize the sociodemographic and locational attributes of the resultant groups. If there exist groups of individuals that are distinctive in both their travel behavior and sociodemographic characteristic, then, for the reasons outlined above, there is a better chance of identifying such groups on the basis of behavior observed over five weeks than on the basis of behavior observed for only one day.

Deriving the travel measures on which to classify individuals

From the information contained in the 35-day travel diaries, one can derive a plethora of measures, each of which describes some aspect of the individual's behavior over the study period. As one cannot use several dozen measures as the basis for classification and as the a priori grounds for selecting a small subset of the key travel variables are not well established, it is necessary to use a data-reduction technique [such as Principal Components Analysis – (PCA)] to identify such a subset. Here we develop an extensive set of travel-activity measures similar to those used in an earlier study (Hanson & Hanson, 1981) but different in a number of important ways because the aims of the two studies are different. We then subject the matrix of travel measures to PCA in order to identify those few travel variables that best summarize an individual's travel-activity pattern and that can then be used for grouping individuals with similar travel behavior.

One of the ways in which longitudinal travel-diary data differ from one-day diaries is the presence in the former of both travel and no-travel days for many individuals. Also, it is worth distinguishing between no-travel days that occur when the person remains at home and no-travel days that occur because the person has left town (e.g., for vacation or a business trip). The distinction is important because in-town no-trip days can be considered days on which the individual is essentially "at risk" to travel, whereas such is not the case for out-of-town days. Perhaps because of its focus on one-day trip records, previous research has had little to say about the importance, or even the existence, of no-trip days. One exception is a study by Brög and Meyburg (1981) who reported that whereas willingness to respond to a (one-day) travel survey has little to do with sociodemographic characteristics, it is very much related to the individual's propensity to travel. Persons unlikely to make a trip on the survey day believe their lack of travel renders them "uninteresting" and "unimportant" to those taking the survey and therefore fail to respond. Furthermore, this effect works differently for personal-interview as opposed to postalquestionnaire travel surveys. In fact the number of days a person did not travel because he or she remained at home may be a measure of travel behavior that is important in identyfying groups of individuals with similar travel-activity patterns. Our analysis takes explicit recognition of both types of no-trip days (in-town and out-of-town no-travel days). The overall measures of trip generation ("number of trips per travel day" and "number of stops per travel day") were based on the days on which the person made at least one trip.

In addition to the number of in-town no-travel days and the number of trips and stops per travel day,[1] many other measures were generated from each individual's longitudinal record. (These are shown in Table 1, which reports the results of the PCA discussed below.) Several variables (the proportions of the individual's trips that were of different stop lengths) were employed to indicate the individual's propensity to make simple as opposed to complex trips. Other measures were the proportions of the individual's stops that were made for each of five different travel purposes (social, personal business, shopping, work, and recreation), by each of five different modes (walk, bike, bus, car driver, car passenger), at each of four different times of day (before 9 a.m., between 9 a.m. and 4 p.m., between 4 p.m. and 7 p.m., and after 7 p.m.), on the weekends as opposed to the week days, and in each of six different sections of the city (within one km of the center of Uppsala; beyond six km of the town center; and, within the remaining five km annulus, the proportion of stops made in each of the four quadrants: southwest, northwest, southeast and northeast). Another set of measures indicated the proportions of the individual's out-of-home time spent in each of the five different purposes. Travel distance measures included "kilometers traveled per travel day" as well as the proportion of the individual's total travel distance associated with pursuing each of the five trip purposes.

We included several measures that are possible to compute only with multiday data; these are indices of the level of variety vs. repetition present in the individual's travel-activity pattern: the proportion of the individual's stops that were made; (1) at different locations (defined as street addresses); (2) at different land use types (identified by a 99-category coding scheme that distinguished, for example, banks from supermarkets from a friend's home); and (3) for different activities (identified by 70 different activity codes that distinguished for instance, "eat meal alone" from "eat meal with others"). To take an example, suppose a person made 200 stops over the 35 days and visited 50 different locations; the "proportion of different locations visited" would be 0.25. These variables measure how dispersed or concentrated the person's stops were in different locations, in different land use types, and in different activity categories. Finally, we included centrographic measures on the point set comprising the spatial distribution of each person's destinations visited during the 35 days (see Hanson & Hanson, 1981, p. 339) as measures of the spatial extent of the individual's travel-activity pattern.

Table 1. Results of principal components analysis of travel measures: Varimax rotated factor loadings.

Factor 1: Shopping and work

Loadings	Variables
.87	Proportion of out-of-home time spent shopping
.81	Proportion of stops made for shopping
.81	Proportion of km traveled for shopping
.77	Proportion of stops made to purchase food
-.73	Proportion of stops made for work
-.72	Proportion of work trips that were single stop
-.67	Proportion of stop that were made before 9 a.m.
-.66	Proportion of out-of-home time spent in work
-.59	Proportion of stops made between 4 p.m. and 7 p.m.
	Percent of variance explained 20.7

Factor 2: Complexity of Travel

Loadings	Variables
-.81	Proportion of trips that were single stop
-.77	Proportion of stops at different locations made on single stop trips
.73	Proportion of trips that had more than 4 stops
.71	Proportion of trips that had 3 stops
.66	Proportion of trips that had 4 stops
	Percent of variance explained 11.5

Factor 3: Frequency of travel

Loadings	Variables
.85	Number of trips per travel day
.78	Number of stops per travel day
-.74	Proportion of stops made for different activities
-.67	Proportion of stops made at different locations
-.53	Number of in-town no-trip days
	Percent of variance explained 9.2

Factor 4: Personal business

Loadings	Variables
.88	Proportion of stops for personal business
.87	Proportion of km travaled for personal business
.84	Proportion of out-of-home time spent in personal business
	Percent of variance explained 8.3

Table 1. Continued.

Factor 5: Spatial extent of activity pattern	
Loadings	Variables
-.78	Proportion of stops made on foot
.76	Distance between home and centroid of activity space
.75	Average distance between home and destinations
.72	Km traveled per travel day
-.52	Proportion of stops within 1.0 km of city center
	Percent of variance explained 6.6
Factor 6: Recreation	
Loadings	Variables
.88	Proportion of stops for recreation
.84	Proportion of km traveled for recreation
.80	Proportion of out-of-home time spent in recreation
	Percent of variance explained 6.3
Factor 7: Social	
Loadings	Variables
.83	Proportion of stops for social activities
.83	Proportion of km traveled for social activities
.73	Proportion of out-of-home time spent in social activities
	Percent of variance explained 4.8
Total percent of variance explained 67.3 by 7 factors	

Although these measures of travel-activity pattern bear close resemblance to the set used by Hanson & Hanson (1981), they differ in several important ways, reflecting the different goals of the two studies. First, the previous analysis took no explicit notice of in-town no-trip days. Second, all travel measures in the previous study (with the exception of the mode use variables) were calculated on the basis of "number of days the individual was at risk to travel in Uppsala" rather than as measures internal to each person's total travel-activity pattern. For example, the variable "number of social stops" was calculated as (number of social stops) ÷ (35 − the number of days the out of town); in this study it was calculated as (number of social stops) ÷ (individual's total number of stops). This reflects the fact that the goal of the previous analysis was primarily to identify the dimensions of travel-activity patterns in the population rather than to define groups of individuals with similar behavior patterns.

Finally, the variables included here relating to time-of-day of travel and to the sectors of Uppsala visited were not included in the earlier study.

We used PCA to reduce the large data matrix to a smaller set of factors, each measuring an "independent" component to be used in grouping individuals with similar travel behavior.[2] The results (Table 1) show that the factors identify the propensity of the individual to undertake travel for each of several purposes (Factors 1, 4, 6, and 7), the overall frequency of travel (Factor 3), the complexity of the individual's trips (Factor 2) and the spatial extent of the individual's travel-activity pattern (Factor 5). These results are interesting in light of the widespread reliance in previous studies on "time spent in different activities" as the dependent variables measuring travel-activity pattern. Our findings support the use of such variables as valid indices of activity pattern, but suggest the need for additional measures to capture overall travel frequency, trip complexity, and destination dispersion.[3]

Clustering individuals on the basis of travel characteristics

The results of the PCA identified the subset of measures that best described parsimoniously an individual's travel-activity pattern and that could be used, therefore, to identify groups of individuals having similar travel characteristics. These measures were the variables with the highest loading on each of the factors with an eigenvalue greater than 1.9.[4] Prior to their use in the clustering routine, some of the variables were transformed so that their distributions more closely approximated the normal distribution. The variables used to cluster individuals were, therefore; (1) the proportion of out-of-home time spent in shopping (square root transformation); (2) the proportion of trips that were single-stop trips; (3) the number of trips per travel day (square root transformation); (4) proportion of stops that were made for personal business (square root transformation); (5) proportion of stops that were made for social purposes (square root transformation); (6) proportion of stops made on foot; and (7) proportions of stops made for recreation (square root transformation).

A k-means clustering algorithm (BMDP, 1981) partitioned the 149 cases into five groups:[5] the procedure yields a pseudo F-ratio indicating the importance of each variable in determining the clusters, and these showed that whereas all seven variables were highly significant discriminators among the groups, the two most important were "proportion of stops made on foot" and "number of trips per travel day."

Table 2, showing the cluster means on each of the seven travel variables, indicates the ways in which the travel behaviors of the individuals in the different clusters vary. In interpreting the clusters, it is useful to keep in mind the dimension of travel that each of the seven variables represents. "Proportion of stops

Table 2. Cluster means on travel measures.

Cluster	Proportion of out-of-home time spent shopping	Proportion of trips that were single stop	Number of trips per traval day	Proportion of stops for personal business	Proportion of stops for social	Proportion of stops on foot	Proportion of stops for recreation
1 N = 11	0.09 (.07)[a]	0.71 (.15)	2.90 (.50)	0.18 (.10)	0.04 (.03)	0.65 (.16)	0.04 (.03)
2 N = 19	0.04 (.04)	0.63 (.18)	2.37 (.34)	0.08 (.07)	0.08 (.06)	0.16 (.09)	0.05 (.03)
3 N = 29	0.04 (.06)	0.73 (.12)	1.50 (0.23)	0.04 (.04)	0.08 (.06)	0.21 (.16)	0.09 (.05)
4 N = 44	0.11 (.08)	0.50 (.18)	1.46 (0.30)	0.08 (.04)	0.09 (.05)	0.72 (.14)	0.07 (.05)
5 N = 46	0.09 (.10)	0.44 (.13)	1.30 (.21)	0.07 (.06)	0.11 (.03)	0.25 (.14)	0.05 (.03)
Total Sample N = 149	0.08 (.08)	0.56 (.19)	1.63 (0.55)	0.08 (.06)	0.09 (.06)	.39 (.27)	0.06 (.04)

[a] (standard deviation).

made on foot", for example, is closely related to other measures of the spatial extent of the travel-activity pattern, and "proportion of trips that were single stop" is tied to other measures of trip complexity (see Table 1). Individuals in Clusters 1 and 2 are set apart by their high levels of trip making and their high proportions of single-stop trips but their relatively low participation in recreational activities. Individuals in Cluster 1, however, spend a distinctly higher proportion of their time in shopping, make an exceptionally large proportion (18 percent) of their stops for personal business (which includes visits to the post office, to a physician's office, to the laundry and so on), make a large proportion (65 percent) of their stops on foot, but make relatively few social stops. Clusters 3, 4, and 5 share lower levels of trip making and relatively high levels of participation in social activities; Cluster 4 stands out as having the highest proportion of time spent in shopping (11 percent) and the largest proportion of pedestrian stops (72 percent). Cluster 3 individuals, on the other hand, make relatively few stops on foot or for personal business or shopping, but make a high proportion of single-stop trips and recreational stops. Members of Cluster 5 make mostly multi-stop trips (only 44 percent of their trips were single purpose), make relatively few stops on foot, and make a high 11 percent of their stops for social purposes.

Discriminating among the groups using sociodemographic and locational variables

If the popular assertion is true that different sociodemographic characteristics are responsible for the development of different travel-activity patterns, then we should be able to associate distinct sociodemographic descriptors with each of these clusters. The individual's residential location vis-a-vis the set of establishments in Uppsala could also play a role in the emergence of one type of travel-activity pattern as opposed to another. The sociodemographic and locational variables shown in Table 3 were used, therefore, to discriminate among the five clusters in a multiple discriminant analysis. The sociodemographics are typical of those used by other researchers who have attempted to relate travel to the characteristics of the individuals and the household. The locational variables in Table 3 were selected from a larger set of locational measures on the basis of results from a previous analysis (Hanson, 1982); they are similar to the measures of zonal density used by Pas (1984) and Recker & Schuler (1982).

The discriminant analysis revealed that two discriminant functions were significant in distinguishing among the five travel clusters:

$D_1 = -1.64$ (sex) $- 0.64$ (household size) $+ 0.0007$ (number of establishments within 1.0 km of home), and

$D_2 =$ 0.36 (household size) + 0.18 (sex) + 0.003 (number of establishments within 1.0 km of home).

Table 3. Sociodemographic and locational variables used as discriminators in multiple discriminant analysis.

Age in Years
Household income, before taxes, in Swedish crowns (sq. root transformation)
Number of people in household
Years of formal education
Number of hours worked per week
Sex
Auto availability: number of autos owned by household ÷ number of persons in household with valid drivers license; index is 0 if person does not hold driver's license or if household owns no cars
Number of establishments within 1.0 km of home
Number of establishments wihtin 4.0 km of home

The eigenvalues associated with each function were 0.44 and 0.14 respectively, indicating that the two functions together account for 58 percent of the variance in the discriminating variables. The canonical correlations, measuring the degree of association between each discriminant function and the five clusters, were 0.55 and 0.35 respectively, indicating that each discriminant function is moderately correlated with the clusters.

Although the same three variables comprise both discriminant functions, the relative importance of sex and household size shifts between the two functions, with sex being the more important in defining the first function, and household size dominant in the second. The density of establishments close to home is statistically significant in discriminating among the clusters, but its relative contribution is notably less than that of the other two variables. The coefficients indicate that persons with high scores on the first discriminant function will be women from smaller households located in areas with many establishments; persons scoring highest on the second function will be men from larger households, also from areas with a high establishment density.

Further insight into the nature of the discriminant functions and the composition of the clusters can be obtained by inspecting Table 4 and the locations of the cluster centroids in the space defined by the two discriminant functions. The first function differentiates Clusters 1, 4, and 5 from Clusters 2 and 3. Although (as seen in Table 4) Clusters 1, 4, and 5 all have a high proportion of women and relatively small households, only Clusters 1 and 4 are located in areas with a high establishment density; the mean number of establishments within 1.0 km of Cluster 5 residences is the lowest for any group. Clusters 2 and 3 differ on the first discriminant function (especially from Clusters 1 and

Table 4. Cluster values on sociodemographic and locational measures.[a]

Cluster	Proportion of cluster that is female	Mean household size	Mean Number of establishments within 1.0 km of home	Proportion of cluster employed full-time	Proportion not employed	Mean Age $\bar{X}$	Mean Auto Avail-ability
1 N = 11	0.80	2.4 (1.1)	451 (407)	0.36	0.36	50.4 (13.2)	0.19 (.25)
2 N = 19	0.21	3.4 (0.7)	278 (389)	0.79	0.10	33.3 (9.7)	0.63 (.32)
3 N = 29	0.31	2.5 (1.0)	265 (288)	0.62	0.14	47.4 (16.5)	0.45 (.43)
4 N = 44	0.75	1.9 (1.1)	480 (366)	0.41	0.46	51.0 (19.3)	0.20 (.33)
5 N = 46	0.67	2.1 (1.0)	180 (234)	0.44	0.41	46.9 (17.1)	0.27 (.43)
Total Sample N = 149	0.57	2.3 (0.4)	315 (321)	0.51	0.32	46.6 (16.7)	0.33 (.38)

[a] Standard deviations are in parentheses.

4) in having a higher proportion of men who come from larger households and lower-density residential environments. The second discriminant function differentiates, rather weakly, Clusters 1, 2, and 4, representing (supposedly) individuals from larger households (but see from Table 4 that Cluster 4 has a very low mean household size), from Cluster 5, a group with fewer people per household living in a low density environment. It is clear that the second discriminant function is not as effective as the first in discriminating among the clusters; it is also clear from perusing Table 4 that the locational variables cannot be relied upon to discriminate consistently among the clusters.

The proximity of the cluster centroids in the space defined by the discriminant functions (especially those of Clusters 1 and 4) suggests that there is considerable overlap among the groups, despite the fact that the discrimination is statistically significant. This is born out by the fact that the discriminant functions could classify only 45 percent of the individuals in the correct cluster. The classification procedure correctly assigned 68 percent of Cluster 2, but only half of Clusters 1 and 5. Individuals in Cluster 4 proved to be the most difficult to classify, with only 29 percent assigned correctly to Cluster 4; about a quarter of the individuals in Cluster 4 were incorrectly assigned to Cluster 1, and a quarter to Cluster 5.

One purpose of the discriminant analysis was to identify those sociodemographic and location variables that best distinguish among the five clusters and that would be therefore the most salient for describing the sociodemographic and locational differences among the groups. The cluster means on each of the three variables identified as the most important are shown in the first three columns of Table 4; the other variables in this Table have been added because their F statistics showed them to be the most important remaining variables and because they help in some cases to interpret the composition of the clusters.

It is now possible to summarize by describing the key characteristics of each cluster.

Cluster 1. Comprised mostly of women from larger households living in areas with a high density of establishments, Cluster 1 is distinguished by a high level of trip making. A large proportion of these trips are single stop, are made on foot, and are made for the purposes of personal business and shopping. Cluster 1 individuals make few of their stops for social or recreational purposes.

Cluster 2. Most members of Cluster 2 are full-time employed males from large households. They are also likely to be young and to have access to a car. Cluster 2 individuals make many trips, but few of these are made on foot. Although very few of their stops are for shopping or recreation a relatively high proportion are to socialize.

Cluster 3. Predominantly men who work full-time, Cluster 3 individuals make relatively fewer trips, but nearly three quarters of these trips have only one stop. Few of their stops are made on foot, for shopping, or for personal business, yet a high proportion are recreational or social stops.

Cluster 4. The typical Cluster 4 individual is a woman living in a small household in an area with a high density of establishments. A relatively large proportion (41 percent) of this cluster is employed full-time, but also a large proportion (46 percent) is not employed. People in Cluster 4 make few trips, about one-half of which have more than one stop. Nearly three quarters of their travel is done on foot, and a high proportion of their time is spent shopping. Social and recreational stops also make up a relatively high proportion of their stops.

Cluster 5. Roughly two thirds women, Cluster 5, like Cluster 4, has high proportions of both full-time employed and not employed people. Members of Cluster 5 come from small households living in areas with relatively few establishments. Although they have the lowest level of trip making of all the clusters, members of this cluster differ from the others in that the majority of their trips (56 percent) have more than one stop. Only a quarter of their stops are made on foot. Although a high proportion of their travel is for shopping and socializing, little is for recreation.

This examination of the travel and personal characteristics of the clusters reveals a number of interesting points. It is worth noting, for example, that there is no consistent relationship between establishment density close to home and the overall level of trip making. This runs counter to the findings of previous studies that have reported an inverse relationship between trip generation and the population density of residential zones (e.g., CATS, 1977); the discrepancy might be explained by the focus of studies like the CATS on vehicular travel and their exclusion of pedestrian trips. Although in our study high establishment density is not linked to low levels of tripmaking, the two groups that make well over half of their stops on foot (Clusters 1 and 4) both live in areas of high establishment density, have low levels of auto availability, and are comprised primarily of women (Table 4). Also, it is interesting to note one respect in which our results are similar to those of Pas (1984); using one-day travel data from Baltimore, he found that (like the members of our Cluster 5) people residing in low-density areas were "...significantly more likely to undertake a multistop daily pattern..." (p. 579). Finally, our results are in tune with those of other (e.g. time budget) studies in that the female-dominated clusters are the ones that spend a high of about 10 percent of their out-of-home time in shopping, whereas the male-dominated groups spend less than half that proportion of their time in that activity.

Discussion

The identification of homogeneous groups defined in terms of similarities in personal traits as well as similarities in travel behavior remains an important practical and theoretical issue in transportation research. Our work to date on this problem suggests that the rather limited success in this regard stems in large part from the mistaken assumption that day-to-day variation in an individual's travel behavior is small relative to the variation between individuals with similar socioeconomic and locational characteristics.

In previous studies individual deviation from the group mean on a measure of travel has been considered "error," but there has been recent recognition that the source of this "error" can probably be specified. Horowitz (1983) has noted the difficulty of determining whether the error represents nonrandom (misspecification) or random (sampling) error. Tardiff (1979) has pointed out that nonrandom error could be attributed to either (1) behaviors that vary among individuals but not among time periods, or (2) behaviors that vary among both individuals and time periods. A question that arises from the results of grouping procedures based on one-day travel data is how much of the variability (e.g., in a travel measure) in a supposedly homogeneous group is due to unexplained heterogeneity within the group (i.e., differences among individuals within a cluster – Tardiff's case (1) above) and how much reflects differences from one day to the next within the individual's longitudinal travel record (i.e., intra-individual differences).

The classification methodology outlined in this paper is in direct response to the conceptual problems of grouping procedures using one-day travel diaries. While recognizing the innate complexity of daily travel behavior, we have "smoothed" out the day-to-day variation in travel at the individual level by using measures of travel derived from 35 days of continuous observation for each individual in the sample. It would be difficult, if not impossible, to compare directly the results of clusters based on multi-day data with clusters based on one-day data because many of the travel measures derived from the 35-day record could not be used with a 1-day record.

It is difficult to compare the results of this study with those of others that have used one-day travel data to identify homogeneous behavioral and sociodemographic groups, mainly because few have attempted to cluster individuals on the basis of several travel variables simultaneously. The more frequently adopted approach is to use as dependent variables several relatively simple measures of activity pattern. Wermuth (1982), for example, used "time spent in activity x" (where x is either work, education, shopping, or recreation) on one survey day as a set of dependent variables; his goal was to account for the variance in each of these a set of sociodemographic and locational descriptors. With such relatively straightforward dependent variables, he finds that personspecific characteristics are able to account for almost all the inter-personal variation in time allocated to work and education; personal and household characteristics explain most of the variance in shopping and recreation, and locational characteristics are relatively unimportant. Allaman *et al.* (1982) use a similar set of time allocation measures (derived from the one-day Baltimore travel diaries), as dependent variables in multiple regressions; their results show employment status, sex, age, and income to be the most important independent variables in explaining variation in time allocation. Using the same data set, Pas (1984) devised a more complex dependent variable by classifying individuals' travel patterns into five types; he then employed the linear logit model to assess the significance of various sociodemographic/locational variables in differentiating among the five patterns and found sex and employment status to be particularly important in that regard.

Our methodology here was quite similar to that used by Recker and Schuler (1982) with the exception that their travel data represented only one day's trips. It is interesting to note the similarities in the discriminant variables that emerged in their study and in ours; the set of variables they found most useful in discriminating among the nine different travel patterns they identified had to do with employment, household roles (closely related to sex), and characteristics of the residential area (employment density and density of single-family dwelling units in the residential area). The two discriminant functions that incorporated these variables were able to classify correctly 39.5 percent of the cases in their sample, somewhat fewer than the 45 percent correctly classified here on the basis of longer term behavior.

Our results do seem to represent a distinct improvement over the attempt by Dangschat *et al.* (1982) to identify groups of individuals that were homogeneous with respect to behavior and sociodemographic characteristics. Although these researchers were able, using data from Germany for one work day and one Sunday, to identify five clusters of individuals with similar within-group behavior patterns (defined as participation in working, shopping, and social activities) they were not able to associate distinctive sociodemographic characteristics with each group: "...in all five clusters the distribution of the sociodemographic characteristics is almost identical with the distribution of the sample" (p. 1172). It is worth noting that our measures of travel behavior encompassed a good deal more behavior complexity than did theirs, and yet our behavior groups were interpretable in terms of sociodemographic and locational variables.

After all is said and done, has the goal of identifying groups of individuals that are homogeneous with respect to both behavior and sociodemographic/locational characteristics been met? Only partially. Using a well established classification methodology on five-week travel records, we have been able to identify five clusters of individuals that share distinctive travel and sociodemographic/locational attributes. There remains, however, considerable intra-group variance with respect to both sets of variables, and there remains substantial overlap among the various groups as demonstrated by the classification procedure's ability to classify correctly only 45 percent of the sample individuals.

We feel that a portion of this unexplained variance could be a function of systematic variation in the daily travel patterns association with any one individual. The travel characteristics associated with each cluster (shown in Table 2) essentially sketch out a *single* travel-activity pattern for each cluster. Yet our evidence suggests that *more than one typical travel pattern* is needed if we wish to effectively characterize the travel behavior of any one individual or any population subgroup for that matter (Huff & Hanson, 1986). This suggests that future classification should recognize the *multiple daily patterns* that an individual and group exhibit. The existence of more than one frequent or typical daily pattern for each individual also raises questions about how we might think about defining habitual behavior. The issues of (1) the level of detail at which travel behavior is measured; (2) the length of time over which travel is measured; (3) what is considered a "habitual" travel pattern; and (4) the travel behavior groups identified in a classification procedure are all very much interrelated. In our view classification procedures in travel research should be based on measures of travel that are sufficiently detailed and complete to distinguish habitual from non-habitual behaviors and that are collected over a long enough period of time to enable habitual behavior patterns to emerge in the data. Grouping procedures should also recognize that each individual has more than one habitual travel pattern.

Notes

1. For example, "number of trips per travel day" was calculated as the total number of trips the person made over the diary period divided by 35 minus the number of no-trip days.

2. We made several PCA runs using different (overlapping) subsets of variables and different (overlapping) subsets of the sample, and the resulting factor structures proved to be extremely robust, with essentially the same factors emerging from each run. Because we wanted to use the PCA results as input to a clustering routine, we used varimax rotation.

3. The factors emerging from this analysis are quite similar to those that emerged in Hanson & Hanson (1981). Both analyses yielded factors describing travel for each of a number of different purposes, a factor describing the spatial extent of the travel-activity pattern, a factor describing travel frequency, and one describing trip complexity. The relative importance of the various factors (as measured by the amount of variance explained) varied between the two sets of results. Another interesting difference was that in the earlier study measures of shopping and trip complexity were combined in one factor, whereas in this study they were not. Here making a high proportion of one's stops for shopping is negatively (but closely) related to making a high proportion of one's stops for work; it is not necessarily related to having a large proportion of multi-stop trips. The

results of the earlier study indicate that those people who made relatively more shopping stops (relative to other people in the sample) were also those who made relatively more multi-stop trips.

4. Factors with lower eigenvalues had only one or two variables with loadings higher than 0.50.

5. We experimented with different numbers of groups and have used the 5-group solution because it was the one that was most interpretable in terms of the sociodemographic variables in Table 3.

References

Adiv, A., 1983. The structure of the work trip based on analysis of trip diaries in the San Francisco bay area. In: S. Carpenter and P. Jones (eds), Recent Advances in Travel Demand Analysis, pp. 117–136. Aldershot, Hants, England. Gower Publishing Company. Ltd.

Allaman, P.M., T.J. Tardiff & F.C. Dunbar, 1982. New Approaches to Understanding Travel Behavior. National Research Council, Washington, D.C.

BMDP Statistical Software 1981. University of California Press, Los Angeles, CA.

Brog, W. & A.H. Meyburg, 1981. The non-response problem in travel surveys: An empirical investigation. Transportation Research Record #775: 34–38.

Chapin, S.S., 1974. Human Activity Pattern in the City. New York: John Wiley and Sons.

Chicago Area Transportation Study (CATS), 1977. A summary of travel characteristics. Chicago: CATS and Northwestern Indians Regional Planning Commission.

Clarke, M.I., M. Dix, P.M. Jones & I.G. Heggie, 1981a. Some recent developments in activity travel analysis and modeling. Transportation Research Record #794, 1–8.

Clarke, M.I., M. Dix & P.M. Jones, 1981b. Error and uncertainty in travel surveys. Transportation 10: 105–126.

Cullen, I. & E. Phelps, 1975. Diary Techniques and the Problem of Urban Life, Final Report to the Social Science Research Council.

Curry, L., 1967. Central places in the random spatial economy. Journal of Regional Science, 7 (supplement): 217–238.

Dangschat, J., W. Droth, J. Friedrichs & K. Kiehl, 1982. Action spaces of urban residents: An empirical study in the region of Hamburg. Environment and Planning A 14: 1155–1174.

Godard, X., 1983. Presentation of some results in the Dijon survey: Linkage between activities and mobility. In: S. Carpenter and P. Jones (eds), Recent Advances in Travel Demand Analysis, pp. 163–181. Aldershot, Hants, England: Gower Publishing Company.

Hanson, S., 1982. The determinants of daily travel activity patterns: Relative location and sociodemographic factors. Urban Geography 3: 179–202.

Hanson, S. & P. Hanson, 1981. The travel activity patterns of urban residents: Dimensions and relationships to sociodemographic characteristics. Economic Geography 57: 332–347.

Hanson, S. & J.O. Huff, 1982. Assessing day to day variability in complex travel patterns. Transportation Research Record #891: 18–24.

Hanson, S. & J.O. Huff, 1986. Repetition and day-to-day variability in individual travel patterns. In: R. Golledge & H. Timmermans (eds), Behavioral Modelling Approaches in Geography and Planning. Reidel. In press.

Herz, R., 1982. The influences of environmental factors on daily behavior. Environment and Planning A 14: 1175–1193.

Herz, R., 1983. Stability, variability, and flexibility in everyday behavior. In: S. Carpenter & P. Jones (eds), Recent Advances in Travel Demand Analysis, pp. 385–400. Aldershot, Hants, England: Gower Publishing Company.

Horowitz, J.L., 1983. Evaluation of discrete-choice, random-utility models as practical tools of transportation systems analysis. Travel Analysis Methods for the 1980s, National Research Council, Washington, D.C.

Huff, J.O. & S. Hanson, 1986. Repetition and variability in urban travel. Geographical Analysis. In press.

Janelle, D.G. & M.F. Goodchild, 1983. Transportation indicators of space-time autonomy. Urban Geography 4: 317–337.

Kutter, E., 1973. A model for individual travel behavior. Urban Studies 10: 235–258.

McCarthy, P.S., 1980. A study of the importance of generalized attributes in shopping choice behavior. Environment and Planning A, 12: 1269–1286.

Oppenheim, N., 1975. A typological approach to individual urban travel behavior prediction. Environment and Planning A, 7: 141–152.

Pas, E., 1984. The effect of selected sociodemographic characteristics on daily travel activity behavior. Environment and Planning A 16: 571–581.

Recker, W. & H. Schuler, 1982. An integrated analysis of complex travel behavior and urban form indicators. Urban Geography 3: 110–120.

Salomon, I., 1983. Lifestyles: A broader perspective on travel behavior. In: S. Carpenter & P. Jones (eds), Recent Advances in Travel Demand Analysis, pp. 290–310. Aldershot, Hants, England.

Szalai, A. (ed.), 1972. The Use of Time. The Hague: Mouton.

Tardiff, T.J., 1979. Definition of alternatives and representation of dynamic behavior in spatial choice models. Transportation Research Record #723: 25–30.

Van der Hoorn, T., 1979. Travel behavior and the total activity pattern. Transportation 8: 309–328.

Vidakovic, V., 1983. The second diary survey in Amsterdam: Some intermediate results. In: S. Carpenter & P. Jones (eds), Recent Demands in Travel Demand Analysis, pp. 172–191. Aldershot, Hants, England. Gower Publishing Co.

Wermuth, M.J., 1982. Hierarchical affects of personal, and residential location characteristics on individual activity demand. Environment and Planning A 14: 1251–1264.